

深度伪造侵权责任承担方式适用困境及完善

董兴佩, 王书平

(山东科技大学 数字法治研究院, 山东 青岛 266590)

摘要:深度伪造侵权案件数量激增反映出传统责任承担机制在应对新型算法侵权时的滞后性。从司法实践各判决来看,非财产责任上,现行停止侵害责任仅能删除已发布的终端视频却无法触及侵权人手中底层人脸生成模型;财产责任上出现赔偿难以量化及不足的现象,这些问题导致被侵权人权利无法得到有效救济且助长深度伪造等新型侵权行为。针对以上问题,应以禁止再次生成伪造内容的方式实现源头治理,将停止侵害的范围扩展至生成模型,平台应利用算法技术向接触过伪造内容的用户精准推送赔礼道歉与恢复名誉的内容,以完善非财产责任承担方式。财产责任上应突破传统损害赔偿理论,借鉴域外不真正无因管理的赔偿逻辑将赔偿范围扩展至侵权人的全部非法经营收益,并适用法定赔偿兜底以及类型化精神损害赔偿的赔偿机制,实现对深度伪造侵权的有效威慑。

关键词:深度伪造;侵权责任承担方式;禁止再次生成伪造内容;无因管理;法定赔偿

中图分类号: D923

文献标识码: A

文章编号: 1672-1217 (2026) 03-0063-09

收稿日期: 2026-02-28

作者简介:董兴佩(1969-),男,山东济宁人,山东科技大学数字法治研究院研究员;

王书平(2001-),女,山东济宁人,山东科技大学文法学院硕士研究生。

DOI:10.16284/j.cnki.cn37-1401/c.2026.03.006

一、问题的提出

深度伪造技术作为生成式人工智能技术的一种表现形式,依托生成对抗网络模型,通过对源数据中特定主体面部、声音及肢体动作等生物特征进行深度学习,进而生成高度逼真的虚假图像、音频或视频内容。^①近年来,我国利用深度伪造技术侵犯个人肖像权、个人信息权益以及信息网络传播权案例逐渐增多,其中对影视作品的换脸还会导致表演者形象的歪曲以及侵害保护作品完整权。^②在这种背景下,司法实践已经对深度伪造侵权这种新型算法侵权作出了回应,以互联网法院审理的AI换脸类案件为例,法院能够查明侵权人是否利用技术手段生成并传播的含有他人肖像的合成图像、视频,并据此认定其行为是否侵害了他人的肖像权或其他人格权益。判决多依据《民法典》的相关规定,判令侵权人删除伪造内容、停止传播并结合案件具体情况确定赔偿责任。^③然而,现有裁判路径按照传统人格权侵权逻辑进行救济,无法有效回应深度伪造可反复生成、侵权成本低特征,这种现象不仅让传统财产责任填平规则无法适用,也造成在实践中非财产救济手段流于表面。因此,亟需针对深度伪造侵权的技术特性,在理论层面重构损害赔偿的计算逻辑,在实践层面优化非财产责任的承担方式,以实现对被侵权人权益的有效救济。

①姚文达、李盼池、赵娅、吴洪超:《人脸深度伪造检测方法研究综述》,《中国图象图形学报》2025年第7期。

②田宏杰:《AI换脸技术的失控风险与法律规制》,《法治社会》2024年第1期。

③河南省许昌市魏都区人民法院(2025)豫1002民初5748号民事判决书;重庆市江北区人民法院(2023)渝0105民初36081号民事判决书;上海市浦东新区人民法院(2023)沪0115民初52178号。

二、深度伪造侵权责任承担方式的适用困境

（一）非财产责任的适用困境

消除影响与赔礼道歉等非财产责任是人格权侵权救济的重要方式，但在面对深度伪造这一新型算法侵权时难以适应侵权行为反复发生的动态特点，救济效果在实践中受到限制。

1. 禁止再次生成伪造内容责任的缺失

我国现行司法实践在处理深度伪造侵权纠纷时适用《民法典》中停止侵害责任，将其理解为对已发布侵权视频的删除。^①然而，这种仅删除伪造内容的责任承担方式忽略了深度伪造技术中生成内容与模型可以分离存在的特点，在生成式人工智能时代侵权风险不再局限于已生成的侵权内容，而是扩展为侵权人掌握的可反复生成的能力。司法裁判停留在删除已发布的伪造内容，未对生成侵权内容的模板进行处理，侵权行为可能在更换账号和平台后再次发生。

相较于我国司法实践侧重于删除伪造内容，美国的监管机构已开始探索对算法生成工具的源头性处理措施。其中，美国联邦贸易委员会在2021年针对Everalbum案的执法行动具有代表性，^②在该案中Everalbum公司未经用户同意的情况下利用用户照片非法训练人脸识别算法，美国联邦贸易委员会在要求其删除原始照片数据的基础上，进一步责令其销毁所有利用非法数据训练生成的算法模型。该案表明，算法模型的训练数据具有违法性，必须从源头销毁以防止侵权行为的持续发生。

深度伪造侵权案与Everalbum案中的侵权风险都不会因删除生成内容而消失，深度伪造侵权行为的制作者未经授权使用肖像或声音数据训练换脸模型，为防止这些换脸模型被反复用于生成新的伪造内容，有必要借鉴源头销毁的执法方式。

2. 赔礼道歉的范围局限性与二次伤害风险

涉及深度伪造侵权的案件中，法院往往会判令侵权人向受害人公开道歉、消除影响并恢复名誉弥补受害人的精神损害，但何时适用、怎样适用这些责任应结合加害行为场景并采取最有助于消除负面影响和恢复受害人名誉的方式。^③在报刊主要传媒时代，消除影响、恢复名誉的常见方式是要求加害人在报刊上登报澄清，这是我国司法裁判主要采取的方式。^④之后进入网络时代，针对网络侵权消除影响、恢复名誉，主要表现为在其网站首页显著位置或者在实施加害行为的网络用户页面作出澄清。

作为高度依赖社交平台算法推荐机制的深度伪造内容的传播，其仅在特定兴趣圈层或通过一对一人机交互链条进行，这样一来在承担公开道歉等非财产责任时就会缺失司法救济的精准度，真正接触过深度伪造内容的特定用户群体会因算法而无法看到发布在用户首页的澄清声明，从未接触过该内容的公众可能会被迫接收澄清信息。司法机关要求侵权人在平台显著位置大范围公开道歉还可能引起史翠珊效应，高曝光度让原本隐秘的伪造内容进入公共视野，舆论二次传播对受害人构成实质性二次伤害。

（二）财产责任的适用困境

由于人格利益本身具有非物质性，在人格权侵权损害赔偿中，无论是财产损害赔偿还是精神损害赔偿，都存在着难以确定数额的问题。但深度伪造侵权传播链条复杂、侵权内容可反复生成、侵权主体多元化的特征使被侵权人实际遭受损失与侵权人所获利益更加难以确定。就精神损害赔偿而言，现有司法解释虽然列举了在确定精神损害赔偿数额时应当考虑的具体因素，但在缺乏客观量化标准的情况下，法官自由裁量空间过大，导致深度伪造侵权中的精神损害赔偿数额认定存在明显差异且赔偿数额整体偏低。

^①杭州互联网法院(2023)浙0192民初4563号民事判决书；成都铁路运输第一法院(2024)川7101民初5615号民事判决书。

^②In the Matter of Everalbum, Inc., Federal Trade Commission, FTC File No. 1923172, January 11, 2021.

^③徐伟：《论生成式人工智能服务提供者侵权损害赔偿》，《财经法学》2025年第2期。

^④河北省易县人民法院(2024)冀0633民初950号民事判决书。

1. 财产损害赔偿难以量化

深度伪造侵权中财产损害赔偿是指由于深度伪造行为导致他人财产损害而发生损害赔偿关系。深度伪造所造成的损害属于人格权益损害,可以分为对不可交易型人格权益的损害和对可交易型人格权益的损害。^①

就可交易型人格权益而言,对于缺乏商业化经验的普通人以及虽为公众人物但未曾将人格权益商业化的群体,在遭遇深度伪造侵权后,因缺乏既往许可交易作为参照,难以准确确定具体受损数额。仅有极少数已形成商业化模式的公众人物,可依据既往许可费构建赔偿基准。法院在处理此类案件时依赖获利返还的处理路径,受制于不当得利的传统法理束缚,难以证明经济损失时由法官酌定赔偿金额,其返还范围被严格限定在现存利益框架内。^②具体到人格权纠纷中,司法裁判习惯性地将侵权人获利视作原本授权许可费,^③这种赔偿计算方式在德国等大陆法系国家的司法实践中也普遍存在,^④但在面对AI换脸等新型侵权时呈现出滞后性。侵权人通过深度伪造内容获取巨额流量广告费甚至带货佣金,而法院仅判赔偿数万元的授权费,二者显然不匹配。这导致侵权人得以变相保留大部分非法收益,遏制深度伪造侵权的威慑作用因此受到削弱。

就不可交易型人格权益而言,例如名誉权、个人信息权益、人格尊严,其本质并不具有市场交易属性,损害也无法通过授权费用、市场价值进行估算。以北京互联网法院审理的侵犯个人信息权益案为例,^⑤由于深度伪造技术替换面部特征使肖像不具有可识别性,法院转而从侵害个人信息权益的角度认定侵权行为。然而相较于肖像权,个人信息权益商业定价与价值评估更为模糊,缺乏统一的赔偿计量标准。司法实践中,法院通常只能依靠酌定来给出一个相对合理数额,在一定程度上弥补损害,但由于缺乏明确且可操作的衡量标准,赔偿数额整体偏低,不足以对深度伪造侵权行为形成威慑,也难以让被侵权人人格利益得到充分修复。

由此可见,无论是可交易型人格权益还是不可交易型人格权益,在深度伪造中均难以通过现行赔偿规则确定赔偿数额,导致《民法典》第1128条确立的损害填平模式在实践适用中受到限制。

2. 精神损害赔偿缺乏明确的标准

精神损害往往具有无形性与难以计量性,《最高人民法院关于确定民事侵权精神损害赔偿责任若干问题的解释》第5条列举了确定精神损害赔偿数额时应当考虑的具体因素,即侵权人过错程度、侵权具体情节、侵权损害后果、侵权人获利情况、侵权人经济能力、当地平均生活水平。该条规定虽然对审判具有一定指导意义,但该规定多为定性描述,缺乏量化标准,在司法实践中精神损害赔偿依赖法官自由裁量,进而导致裁判结果差异大。

在司法实践中,涉及侮辱、丑化、误导公众造成社会评价降低或明显精神困扰时法院才倾向于支持精神损害赔偿。^⑥在支持赔偿的情况下,法院通常判决数千元至两三万元的精神损害抚慰金。对于技术开发成本低廉、收益巨大的深度伪造侵权产业来说,难以形成有效威慑。以个人信息权益侵害、肖像权侵害为裁判基础时,则出现同案不同判情形。在吴某诉某公司网络责任侵权纠纷一案中,^⑦侧重于权利受损的风险性,侵权人侵害被侵权人对于个人信息的处理决定权,使被侵权人的个人信息置于被滥用、泄露的风险中,尽管一般公众无法准确识别到被侵权人本人,损害后果较轻,法官仍酌情支持500元的精神抚慰金。而在殷某某诉北京某智能科技公司人格权纠纷案中,^⑧法院则严格按照严重后果的标准,以殷某某并未举证证明

① 徐银波:《侵权损害赔偿论》,北京:中国法制出版社,2014年,第51页。

② 金鑫、潘振飞、周温涛:《AI换脸中肖像权的延伸保护》,《人民司法》2024年第31期。

③ 北京市第一中级人民法院(2018)京01民终97号民事判决书。

④ [德]维克托莉亚·克雷齐格:《针对人工智能生成影像侵权的保护》,邓环宇译,《南大法学》2025年第2期。

⑤ 北京互联网法院(2023)京0491民初3821号民事判决书。

⑥ 北京互联网法院发布“涉人工智能典型案例”之案例六:程某诉孙某网络侵权责任纠纷案。

⑦ 北京互联网法院(2023)京0491民初3821号民事判决书。

⑧ 北京互联网法院(2023)京0491民初12142号民事判决书。

被诉侵权行为给其精神利益造成严重损害为由，驳回其精神损害赔偿请求。因此，可以看出在深度伪造侵权中是否支持精神损害赔偿尚未形成统一共识，反映出现行侵权规则在应对此类新型侵权时仍存在不足。

三、深度伪造侵权责任承担适用困境的成因

深度伪造侵权责任的承担面临着诸多适用困境，这些困境的产生并非偶然，而是由多方面因素共同作用导致的，深入剖析其成因对于寻求合理的解决路径至关重要。

（一）填平原则失灵

我国现行侵权责任体系所遵循的使受害人利益恢复到未受损害前的状态的填平规则，在处理交通事故、财产毁损等传统纠纷时因可通过常规损害填补方式实现救济而显现出有效性，但在面对具有显著的低成本、高收益特征的深度伪造侵权时就陷入失灵的困境。过去制作好莱坞级别特效换脸视频需要专业团队、昂贵设备和长期制作时间，如今利用各种社交媒体和换脸 APP，普通用户仅需几分钟且花费极少甚至免费，就可生成高度逼真视频与图像，譬如用于干扰美国新罕布什尔州初选的拜登总统伪造录音，制作成本仅为 1 美元且耗时不到 20 分钟。^① 侵权者可以利用一键生成技术批量生产伪造内容，短时间内针对特定目标生成成千上万个色情、诈骗视频，这种广撒网以及低成本方式的侵权方式，让侵权人通过网络传播获取流量收益、广告分成和诈骗所得数额极高，然而受害人实际损失，无论是精神抚慰金还是许可费都难以量化且数额较小。当侵权收益远高于法定赔偿成本时，侵权人往往将法律赔偿视为可承受的经营成本。一旦获利足以覆盖赔偿，法律不仅难以发挥遏制作用，反而变相鼓励了侵权行为持续发生。

（二）非财产责任承担方式难以与动态算法匹配

深度伪造技术具有反复生成的特性，而现有责任承担方式即停止侵害、消除影响是建立在删除即停止的传统侵权责任承担认知之上，譬如在互联网早期，删除侵权帖文通常意味着传播中断，生成式 AI 时代的侵权内容可以云储存且以极低成本反复生成，司法判决通常只能针对特定视频链接和特定平台下达禁令。被侵权人通过诉讼删除了一条视频，侵权人还可以在几秒钟内生成并在其他平台继续发布，传统停止侵害责任承担方式无法实质性制止侵权的持续发生。消除危险责任承担方式是通过预防性救济手段阻断侵权风险持续或再次发生，然而深度伪造技术下该责任形态实际适用面临着技术层面障碍，深度学习的模型参数与训练数据之间并不存在直观、可追溯的对应关系，无论是被侵权人还是司法机关均难以确认侵权人是否已真正删除侵权信息。

现有法律救济往往盯着已经发布出来的视频内容却忽视技术模型的反复生成性，从技术原理来看，深度伪造的本质是将受害人的生物识别特征转化为代码，并内嵌于生成式模型之中，司法裁判仅止步于移除已发布的表层视频无异于采取了一种治标不治本的形式主义，只要技术源头未被切断，侵权人就保留低成本、大规模再次生成侵权内容的能力成为侵权隐患。

（三）缺少商业价值的参照与多重权利竞合

传统认知中只有公众人物肖像才具有明确商业价值，普通人人格权益通常被视为精神利益载体且不具备可交易性，然而深度伪造技术的出现彻底改变这一格局，深度伪造技术能够低成本地将任何人的面部、声音植入到色情、诈骗或广告视频中进行牟利。这使普通人的人格要素被纳入商业利用的范围，由于普通人缺乏明星般的代言对价或出场费作为参照标准，判决时因缺乏量化依据而陷入低赔偿的裁量困境，现有精神抚慰金制度仅具有有限的补偿功能，数额远不足以覆盖侵权人利用深度伪造技术获取的暴利，司法实践中就出现了要么没法赔、要么赔得少的尴尬局面。人格利益的商业化导致损害数额难以精确评估，深度伪造技术本身则加剧了侵权行为的定性困难。以制作深度伪造诈骗视频为例，侵权行为擅自使用被侵权人的肖像与声音，不仅构成对肖像权和声音权的侵害，还在技术处理过程中利用被侵权人面部特征、声纹等

^①David Ingram, "I Created the Biden Deepfake: Magician Says He Was Paid \$1,500 by Steve Kramer to Make the Robo-call," *NBC News*, February 23, 2024.

生物识别信息,涉及个人信息权益的非法处理,相关虚假内容亦可能损害受害人的社会评价,甚至触及其私人生活领域,侵害其名誉权和隐私权。多个人格权益在同一行为中同时受损,在具体案件中难以划分各项权利的受损范围,也增加了责任认定中重复评价或遗漏的风险。

(四)技术特征导致侵权主体与因果关系认定困难

赔偿标准缺少统一性主要体现为责任承担中的量化困难,在深度伪造侵权中更为棘手的问题是责任认定本身的定性困难,深度伪造并非由单一主体独立实施,其侵权后果往往是在多重行为连续作用下逐步放大的结果,呈现出明显的多因一果特征。比如,深度伪造视频虽然侵犯明星肖像权但同时也带来巨大流量关注,如何根据相关证据精准计算由伪造行为导致的直接损失是一大难题。在缺乏统一量化标准的情况下,不同主体在侵权行为中分工交错、功能叠加,使得侵权行为难以简单归属于单一行为人,^①法官难以对不同环节在损害形成中的作用进行精确区分,更难以量化各主体对最终损害后果所作出的贡献程度,即便受害人遭受严重的人格损害,也难以通过现有规则获得与损害程度相匹配的赔偿救济。

深度伪造侵权案件中若严格依循“谁主张,谁举证”一般原则,要求受害人对涉案视听内容伪造性承担证明责任,意味着其必须依赖专业技术鉴定来完成举证,现实中此类鉴定成本高昂且原始文件及其相关数据均由平台或内容发布者掌握,要求当事人证明某一内容为深度伪造,难度明显超出普通诉讼主体举证能力范围。反之,采取举证责任倒置方式转而要求内容发布者证明相关视频真实,能解决受害人证明的困境,但可能引发大量恶意诉讼,也会对正常内容传播形成过度限制,影响言论自由。

四、深度伪造侵权责任承担方式的完善

深度伪造侵权的特殊性使得传统侵权责任承担在适用上产生问题,若不能妥善解决既会导致受害人权利无法得到有效保障,也会助长深度伪造等新型侵权行为,因而需要进一步优化完善非财产责任与财产责任,以更好地保护被侵权人的权利。

(一)优化非财产责任的承担方式

1.确立禁止再次生成伪造内容的责任

与传统侵权不同的是深度伪造技术是利用数据模型可以反复实施侵权,侵权人一旦掌握已训练成熟的生成模型,便能以低成本低地反复实施侵权行为,再次侵权风险显著高于传统侵权。具体到个案裁决层面,判决不能仅以删除已生成内容作为停止侵害的实现方式,应依据技术评估得出再生成几率,^②在判决主文中明确加入禁止再次生成的特定责任承担方式。以往司法实践中,如 AI 换脸公益诉讼案,其判决内容往往止步于责令删除已发布的终端内容,对底层模型的删除未作出说明,^③为再次生成侵权内容留下空间。为实现源头治理,我国司法裁判应当借鉴美国联邦贸易委员会在 Everalbum 案中确立的源头销毁的逻辑,不采纳侵权人技术中立的抗辩。究其根本,用于侵权的 AI 换脸技术并非完全中立的工具,而是带有伪造的特殊属性。^④在我国法律框架之下,可引入禁止再次生成责任,将《民法典》中停止侵害、消除危险责任形式在 AI 技术条件下具体化,在判决主文中明确加入禁止再次生成伪造内容的特定责任,责令侵权人彻底删除云端及本地的换脸模板与底层数据。

禁止再次生成责任落地,必须建立一套科学、可操作的判定标准,不同类型模型之间责任承担方式应区分适用。就主要功能是模仿特定主体、侵权倾向高度集中的专用模型而言,应当通过删除模型、技术阻断等方式从源头禁止其再次生成;对于具有一定合法用途但具有侵权可能性的模型,则通过删除相关数据、重新训练等手段降低其再次侵权风险。而通用基础模型,优先通过关键词过滤等方式防止侵权再次发生。

^①彭桂兵、张风逸:《“深度伪造”侵犯肖像权问题研究》,《湖南师范大学社会科学学报》2023 年第 6 期。

^②林涸民:《论人工智能立法的基本路径》,《中国法学》2024 年第 5 期。

^③杭州互联网法院(2023)浙 0192 民初 4563 号民事判决书。

^④林爱珺、林倩敏:《AI 换脸的技术风险与多元规制》,《未来传播》2023 年第 1 期。

2. 完善赔礼道歉与恢复名誉的范围与方式

深度伪造侵权案件中要使赔礼道歉与恢复名誉真正起到作用,要向最有可能接触到侵权内容的主体发布澄清声明。侵权内容通常依托具体平台或社交媒体进行传播,最能够有效消除影响、恢复受害人名誉的方式,应当是由相关平台主动向曾接触过该侵权内容的用户推送澄清声明。然而,此类个性化告知在技术层面是否可行、平台运营成本是否能够承受,均需纳入制度设计考量。如果平台以及社交媒体能够以合理成本识别出在一定期限内曾观看、转发或下载相同或相似深度伪造内容的用户,那么向这些用户推送澄清声明,显然是最为精准且效果最佳的消除影响方式。多数内容平台为维持服务质量通常会留存一定期限内的内容访问记录,仅从技术层面而言,平台具备判断何人曾接触过特定内容并向其定向推送澄清信息的能力;就实践案例而言,早在2017年微信就在治理网络谣言中应用了精准辟谣机制,当信息被确认为是虚假的,平台系统会自动向所有曾经阅读、点赞和转发过该信息的用户发送定向的系统通知,告知其内容不实。^①就运营成本而言,如今各平台都高度依赖算法推荐,基于用户浏览内容的偏好进行用户画像以实现广告的精准投送,将此种定向投送的技术用于侵权澄清声明的推送成本是较低的。因此,要求平台承担此项责任既不会造成不合理的成本负担,也不违背平台运营规律,是当下最精准且效果最佳的消除影响的方式。

有些平台会因技术困难或者隐私条款无法精准触达受众,例如平台基于隐私保护义务不得随意查阅用户浏览日志,公告澄清作为一种替代性的救济路径在此时可以发挥作用,既然点对点的精准告知难以实现,平台就要承担起公示义务,在应用程序的显著位置设立专门的澄清板块。可参照电子商务领域的监管规范要求电商平台,设置专门区域公示知识产权侵权通知与经营者声明,以解决假冒伪劣商品信息不对称问题。^②深度伪造内容服务提供者也可以设置这种侵权事实澄清专区,集中展示对深度伪造内容的声明,有效阻断虚假信息的传播并最大限度修复受害人名誉,当然也可作为平台已履行合理注意义务的证明。

(二) 完善财产责任的承担方式

针对深度伪造侵权中难以量化的困境,有必要构建多层次的救济体系。实践适用时必须理顺本文提出的各类赔偿方式与传统侵权损害赔偿之间的顺位关系,被侵权人实际损失难以证明时,应优先适用本文引入的不法管理获利返还作为替代。实际损失与侵权获利均无法查明时,启用法定赔偿作为兜底。三者只能择一适用。而类型化精神损害赔偿目的是弥补人格尊严的受损,惩罚性赔偿旨在惩戒恶意侵权,二者互不排斥,且可以与法定赔偿并行适用。

1. 引入不真正无因管理的赔偿逻辑

我国现行的侵权责任对人格权的保护主要遵循填平原则,赔偿范围局限于相当于授权费的损失且权利人需对侵权人的主观过错承担举证责任,在深度伪造侵权中制度局限性日益凸显。相比之下,瑞士民法典在法律人格保护第28a部分规定,原告得向法官提出以下诉讼:有受侵害之虞时,得提起防止侵害的诉讼;对已发生的侵害,得提起除去侵害的诉讼。^③瑞士债务法第423条规定:“管理人,非为本人利益而管理事务者,本人仍有权取得因事务管理而产生的利益。对于管理人所受之损害及所支出之费用,本人仅以其所得利益为限负返还义务。”^④该制度赋予被侵权人主张剥夺侵权人因擅自管理其人格要素而获得的全部不法收益的权利,包括广告费、流量分成等,突破传统侵权法以填平原则的局限,不再以虚拟授权费作为赔偿数额的参考,将其扩展至涵盖肖像、姓名等被商业化利用后的所有经营性利润,在构成要件上也不需要主观过错的举证,解决了在深度伪造案件中因损失难以量化而导致的低赔偿困境。只要证明行为人对受害人的肖像、声音等具财产属性的人格权益实施了无因管理行为,就可启动全部获利返还机制。^⑤

^① 张修智:《“辟谣骑士”的风车大战》,《新华每日电讯》2017年12月15日。

^② 徐伟:《论生成式人工智能服务提供者侵权损害赔偿》,《财经法学》2025年第2期。

^③ 《瑞士民法典》,戴永盛译,北京:中国政法大学出版社,2016年,第13页。

^④ 《瑞士债务法》,戴永盛译,北京:中国政法大学出版社,2016年,第202页。

^⑤ 吕纯顺:《AI换脸侵权问题研究》,《上海法学研究集刊》2023年第6辑。

首先,要突破传统无因管理中的管理意思的主观要件束缚,直接适用不真正无因管理的判定逻辑。传统无因管理要求管理人具有为他人管理事务的主观意思,如果强行用在深度伪造侵权中,显然在主观上存在法理的逻辑冲突,解决这一矛盾的关键在于引入不真正无因管理。在深度伪造侵权中,尽管侵权人主观上是为自己牟利,但由于其擅自干涉归属于他人的人格权益,客观上使本该归属于被侵权人的利益发生了转移,这完全契合不真正无因管理中的不法管理的构成要件。基于任何人不得因不法行为获利原则,法律授权被侵权人可以主张将侵权人利用人格利益经营的行为视为是替自己进行的,要求剥夺其全部非法收益。其次,应严格限定适用条件,区分善意与恶意。不法管理实质上带有一种民事惩罚性质,因此必须严格限制其适用范围,避免过分干涉商业自由。建议明确该规则仅适用于具有主观恶意的侵权人,对于善意或过失的侵权人,如未尽审核义务的平台和不知情的转发者,仍适用传统的损害赔偿规则,以填补损失和返还现存利益为限;仅对于明知是他人肖像和声音而故意进行商业化利用的恶意侵权人,如专门制作换脸视频牟利的制作者,才适用该规则。

2.类型化精神损害赔偿

根据统计数据可知,在人格权纠纷中法院支持适用精神损害赔偿的案件在25%~40%之间,而且大部分是侵权行为造成当事人严重的精神损害。^①这在一定程度上反映出司法实践中对精神损害赔偿持审慎态度,同时精神损害赔偿数额标准不统一,出现同案不同判的现象。与财产损害不同的是,精神损害本身具有主观性,其程度难以通过客观指标进行量化。为避免精神损害赔偿过度依赖法官自由裁量,建议借鉴美国法的分级思维,专门构建一套针对精神抚慰金的类型化思路。美国《深度伪造侵权责任法案》设定的赔偿机制,根据侵权内容对受害人名誉、尊严的损害程度,独立计算精神赔偿数额,该数额可与前述财产损害赔偿叠加适用。受害人可请求侵权人赔偿实际损失及交付因信息披露违法所得;难以精确核算时可主张更高的法定赔偿。法定赔偿按恶性程度分为三档:造成实际损害或存在重大风险的,最高不超过5万美元;已致重大实际损害且不停止并继续捏造、篡改、扩散的,最高不超过10万美元;以羞辱或骚扰为目的、含明显色情内容、虚构特定个人性行为并严重贬损社会评价的,最高不超过15万美元。^②

我国可以借鉴以深度伪造产物的内容性质为标准,对精神损害赔偿的赔偿标准进行类型化界定。具体来讲,可以将侵权者的侵权行为分为以下三类:一是深度伪造产物不涉及侮辱性或非谤性内容,深度伪造产物并没有在网络平台大面积传播,例如侵权人为了娱乐将影视明星的面部肖像替换为自己或他人的肖像,此类情况对权利人造成的精神创伤较小,可以按照最低的赔偿标准进行赔偿。二是深度伪造合成物涉及侮辱性或非谤性内容,例如对公众人物的面部肖像进行恶搞处理丑化其形象,或是对特定权利人的声音进行恶意变声,将其正常的音色修改为容易引起他人不适的另类音色。在此类情况下,可以提升精神损害赔偿的赔偿数额。三是深度伪造合成物涉及色情、暴力、性骚扰等内容,例如侵权人将女明星的面部形象替换到其他主体录制的色情视频中,此类涉及色情、暴力、性骚扰等内容的侵权行为对受害人会产生重大不良影响,受害人可以申请最高数额范围的精神损害赔偿。此外,若深度伪造内容对被侵权人造成严重精神创伤,进而引发抑郁、创伤后应激障碍等心理疾病,因此发生的心理治疗费用应纳入赔偿范围。最高人民法院在涉及未成年人遭受性侵害的司法解释中明确指出,未成年被害人为治疗心理创伤所需的费用可以依法获得支持,并作为物质损失的一部分。这一做法体现对人格尊严与精神健康的保护,值得在人格权侵权案件中加以借鉴。被侵权人提出该赔偿请求时,法院可以要求其提供相关的鉴定意见或者医疗诊断证明,来证明心理治疗费用赔偿的必要性。

3.引入惩罚性赔偿

在虞某某案中,行为人在未获得授权的情况下,利用AI换脸软件将从互联网等渠道收集到的他人人

^①吴睿正:《人脸深度合成技术背景下〈民法典〉第1019条的适用困境及其纾解——基于23份民事判决书的实证分析》,载北京大学人口研究所等编:《首届AI+社会治理·青年学术论坛论文集》,北京:北京大学人口研究所,2025年,第253-262页。

^②DEEPFAKES Accountability Act, H.R. 5586, 118th Cong. § 2 (2023). (Proposed Bill).

脸信息与部分淫秽视频中的人脸信息进行替换合成,制作生成虚假的换脸视频,在网络社交软件上进行广泛传播。^①该案件中侵权行为造成的社会危害后果极其严重,但受害人难以举证证明具体的财产损失。在这种情况下,单纯填平原则的赔偿远低于侵权人的获利,导致赔偿缺乏法律威慑力。虽然我国《民法典》人格权编尚未明文规定惩罚性赔偿,但在知识产权、产品责任以及环境污染中均确立了惩罚性赔偿,证明立法者在这种风险高、隐蔽性强并且难举证的侵权行为中选择更具有威慑性的责任方式。因此,对于情节恶劣并造成严重损害后果的深度伪造侵权,可考虑在人格权侵权中引入惩罚性赔偿制度。

深度伪造侵权与上述侵权在规模、证明责任、侵害社会公共利益方面具有高度相似性,且深度伪造侵权技术提供者通常是大型商业主体,赔偿金被视为可以接受的经营成本。深度伪造的大规模侵权中,涉及被侵权人数量多,但并非所有被侵权人都能发现侵权行为且愿意提起诉讼,实际承担赔偿责任的概率偏低。当侵权收益大于违法成本时,侵权人就会铤而走险实施侵权行为。因此有必要引入惩罚性赔偿,让侵权人承担超出实际损失的赔偿责任,提高其违法成本,实现对潜在侵权行为的有效遏制。

惩罚性赔偿适用要以造成严重后果为前置条件,但“严重”一词本身就是一个开放性的法律概念,赋予法官充分的自由裁量权。在判断知识产权侵权是否构成情节严重时,列举可用于认定情节严重的多种具体行为。具体可以包括侵权次数、持续时间、损害结果、获利大小、是否损害公共利益、此前是否接受过处罚等,帮助法院更好地进行衡量和判定。是否存在组织化或产业化的黑产运营,是否存在规避深度合成服务规范的违规行为,内容是否涉及未成年人、色情利用或侮辱性信息,以及侵权人是否具有主观故意或重大过失,可以作为深度伪造侵权是否为情节严重的考量因素,具体赔偿幅度应构建与其主观恶性及客观后果相适配的阶梯式规则。未造成严重后果的一般侵权应排除惩罚性赔偿的适用,对具有侮辱性、屡教不改或攫取巨额非法利益并引致严重后果的,可设定为受害人精神损害赔偿数额的五倍以下,针对涉及色情内容、侵害未成年人权益,造成受害人严重心理创伤及广泛社会负面影响的极端恶劣情形,建议将惩罚性赔偿上限提升至精神损害赔偿的十倍。

4. 引入法定赔偿规则

针对财产责任救济中损失难以证明的困境,有必要引入法定赔偿规则。在被侵权人实际财产损失或侵权人违法获利均无法通过证据确切计算,以及不可交易型人格权财产损害赔偿无法确定的情况下,法院可直接适用法定赔偿标准进行兜底,以解决深度伪造中人格权财产价值难以量化的问题。对此,个人信息保护领域的立法经验可提供参照价值,在个人信息权益侵权中同样面临被侵权人财产损失金额难以判断的问题,为此,我国台湾地区“个人资料保护法”第28条第3款采用每人每一事件新台币五百元以上二万元以下的法定赔偿标准。在《个人信息保护法》立法过程中,有学者建议按每起侵权事件向每位受害人赔偿500元至1000元不等,但立法机关最终未采纳该建议,原因在于个人信息侵害案件往往涉及人数众多,若采用统一金额的法定赔偿可能导致个人信息处理者承担过高的整体赔偿额。^②

不同于个人信息处理行为中常见的大规模、无差别收集,深度伪造侵权生成内容通常具有针对性,其技术路径多系围绕特定自然人的影像、声音等素材展开,并非同时侵害海量非特定主体的人格利益。实践中即使出现大规模传播的深度伪造内容,也多是对同一受害人影像重复使用,并非同时侵害大量不同主体的人格利益。整体来看,深度伪造侵权中设立一定幅度的法定赔偿标准能够避免因损害举证困难而无法获得赔偿的情形,同时也不会对内容平台或制作者造成难以承受的赔偿负担。

法定赔偿真正落地依赖于通过司法解释确定它的适用顺位,只有在实际损失和获利无法查明时才能启用这一兜底性救济,且其适用不排除精神损害抚慰金并行主张,额度设计可以参照知识产权法划定的弹性区间,基于侵权行为的性质、主观恶意及损害后果建议将深度伪造侵权的法定赔偿划分为四个层级。其一为轻微侵权,针对个人非商业目的技术测试或无意泄露,设定一千元至一万元的赔偿区间;其二为一般商业侵权,针对未经授权将他人人格要素用于商业场景,赔偿额度提升至一万元至五万元;其三为严重侵权,

^①杭州互联网法院(2023)浙0192民初4563号民事判决书。

^②程啸:《个人信息保护法理解与适用》,北京:中国法制出版社,2021年,第521页。

涉及深度伪造色情、诈骗或侮辱诽谤内容的，应适用五万元至十万元的赔偿标准；其四为情节特别严重，对于以制作非法伪造内容为核心业态、呈现产业化与组织化特征且拒不整改的黑产行为，适用十万元至二十万元的高额赔偿，以实现惩罚性威慑。

综上所述，面对深度伪造侵权此类新型算法侵权，传统救济体系亟需突破现有责任承担困境。就非财产责任而言，难以有效应对深度伪造内容可反复生成的风险，除完善赔礼道歉与恢复名誉的方式外，应引入禁止再次生成伪造内容的责任承担方式，阻断生成伪造内容的源头。就财产责任而言，针对损害赔偿难以量化以及数额过低的问题，可借鉴域外制度，尝试通过不法管理的逻辑优化获利返还机制，将精神损害赔偿类型化，引入法定赔偿进行兜底。在情节严重的侵权情形下，补偿性赔偿责任不足以形成威慑，有必要在严格限定前提条件下适用惩罚性赔偿以提高违法成本。

Difficulties in the Application of Liability Remedies for Deepfake Infringement and Proposals for Improvement

DONG Xing-pei, WANG Shu-ping

(Institute of Digital Rule of Law, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: The rapid increase in deepfake infringement cases reflects the lag of traditional liability mechanisms in responding to novel forms of algorithmic torts. Judicial practice indicates that, with respect to non-pecuniary liability, the current duty to cease infringement can only require the deletion of already published videos but cannot reach the underlying facial-generation models possessed by the infringer. Regarding pecuniary liability, courts often face difficulties in quantifying damages, and the resulting compensation is frequently insufficient. These problems prevent victims from obtaining effective remedies and further encourage deepfakes and other emerging forms of infringement. To address these issues, governance should focus on source control by prohibiting the further generation of forged content and by extending the scope of the duty to cease infringement to generative models. In addition, platforms should utilize algorithmic technologies to precisely deliver apology statements and reputation-restoring information to users who have been exposed to the fabricated content, thereby improving the mechanisms for non-pecuniary liability. With regard to pecuniary liability, it is necessary to move beyond the traditional theory of damages. Drawing on the compensation logic of unauthentic negotiorum gestio from foreign jurisdictions, the scope of compensation should be expanded to cover all illicit operating profits obtained by the infringer. At the same time, statutory damages should serve as a fallback mechanism, supplemented by a categorized system of compensation for mental distress, in order to achieve effective deterrence against deepfake-related infringements.

Keywords: deepfake; liability remedies for infringement; prohibition of repeated generation of deepfake content; negotiorum gestio; statutory damages

[责任编辑 治平]